Marcin Śmiałek

AGENTS: WHEN THINGS GO WRONG

Plan of presentation

- Introduction
- Electric Elves – agents in society
  - Genuine Vs. Artificial
- A Security Model for Multi-Agent Systems
  - Agents environments are dynamic, but is security?
- Franco and Agents' Hell
  - A scenario of worst practices

Introduction

- Agents are rarely the best solution.
- Murphy’s Law does work.
- Think about users.
- There are social problems.
- Security policy regulations work against agents.

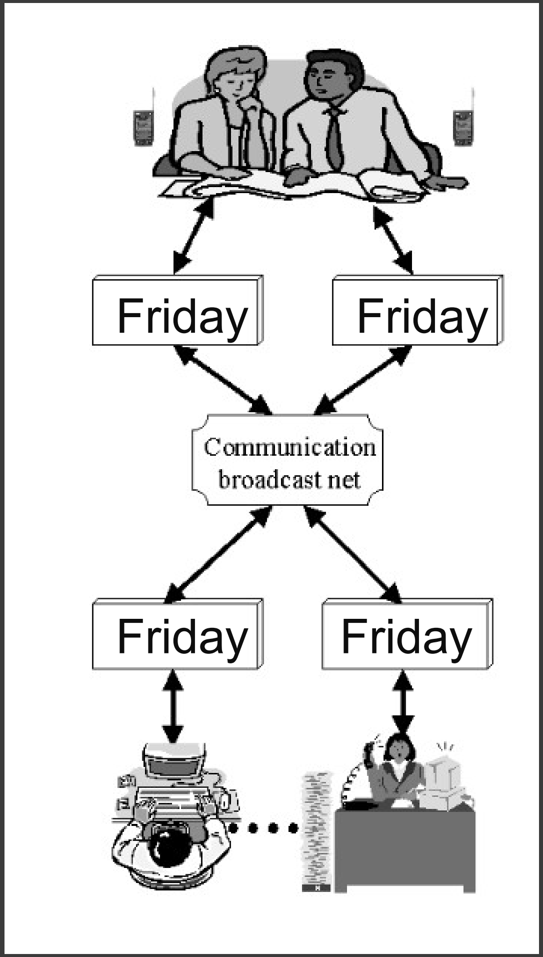
Electric Elves

- Software personal assistants of different kinds.
- Office environment.
- Privacy
- Adjustable autonomy
- Social norms

Goals of the project

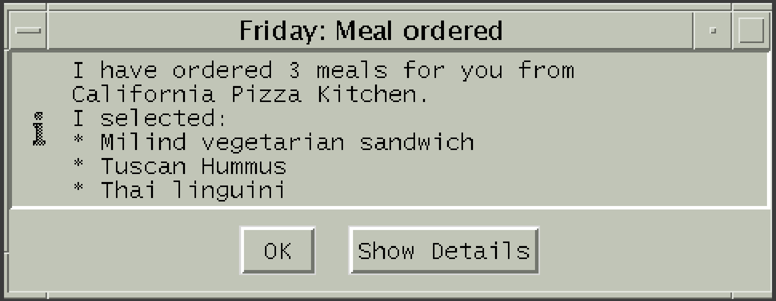
- Reduce the burden on humans.
- Help with organization of time and resources.
- Monitoring of activities.
- Autonomous decisions.
- Communication proxy.
- Utilization of stationary and mobile devices.

Architecture and devices

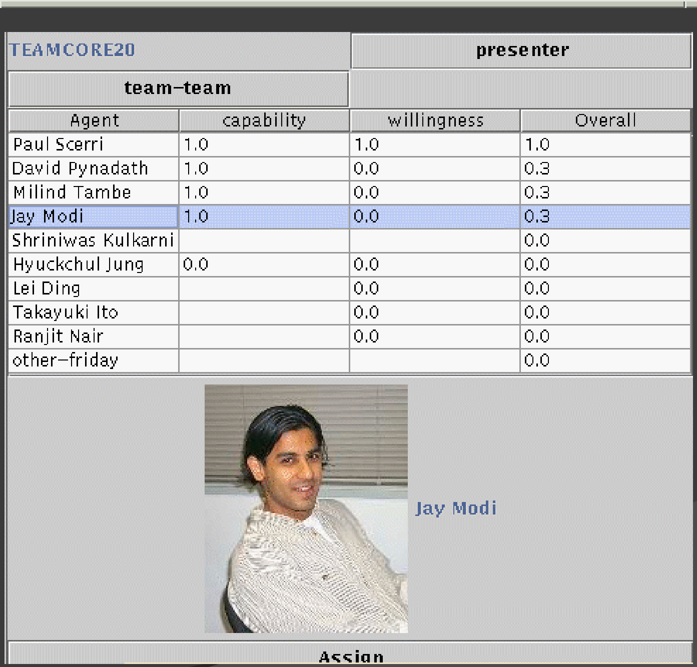


Interface and decisions

- In case you don’t see pictures:
  - Food order notification
  - New presenter election
- Program can decide or ask the user



```
I have ordered 3 meals for you from
California Pizza Kitchen.
I selected:
* Milind vegetarian sandwich
* Tuscan Hummus
* Thai linguini
```



Agent	capability	willingness	Overall
Paul Scerri	1.0	1.0	1.0
David Pynadath	1.0	0.0	0.3
Milind Tambe	1.0	0.0	0.3
Jay Modi	1.0	0.0	0.3
Shriniwas Kulkarni			0.0
Hyuckchul Jung	0.0	0.0	0.0
Lei Ding		0.0	0.0
Takayuki Ito		0.0	0.0
Ranjit Nair		0.0	0.0
other-friday			0.0

Adjustable Autonomy

- Make autonomous decision…
- Transfer control…
- Change coordination constraints.
- Some decisions may be costly or even dangerous.
- Decision-tree learning.
- Ask user if reasoning is proper.

The decision rules.

- IF *two person meeting with important person AND user not at department at meeting time THEN delay the meeting 15 minutes.*
- It’s too problematic for the user to manually input rules.
- System has to learn.

Examples of failures.

- A Friday autonomously cancelled a meeting with the division director because Friday over-generalized from training examples.
- A Friday incorrectly cancelled the group's weekly research meeting when a time-out forced the choice of an autonomous action when user did not respond.

Examples continued.

- A Friday delayed a meeting almost 50 times, each time by 5 minutes. It was correctly applying a learned rule but ignoring the nuisance to the rest of the meeting participants.
- Tambe’s (one of users) Friday automatically volunteered him for a presentation, but he was actually unwilling. Again Friday had over-generalized from a few examples and when a timeout occurred had taken an undesirable autonomous action.

Possible solutions

- Avoid risky decisions – buy user more time.
- Deal with failures of the user to respond.
- Plan ahead to avoid costly sequences.
- Decision trees with big amount of data.
- “Smart” agents dealing with uncertainty and sensing problems.
- Expectable agents.

Privacy

- Privacy was not considered important.
- Software assistants lead to privacy lost.
- Users may feel uncomfortable, what leads to decrease of efficiency.

Privacy and social problems.

- They used GPS to check if coworker is nearby, to arrange a meeting.
- Information availability.
- Conflicts between employees.
- There are no satisfying solutions.
- Is lying good?
- Manipulation - abuse of agent “autonomy”.

A Security Model for Agents

- Security with excessive restrictions means no security at all.
- Information and code must be shared across networks of system without any common administrative control.
- Control of the protection system must be delegated.
- Security must be correct.

Current solutions are bad.

- Identifying the principal is not obvious.
- Atomic view of principal is not for agents.
- Design permissions not for user but for cooperation.
- Remember about buffer overrun.
- FIPA Security tells much about protocols, not the model.

Ideas that might work.

- Principal-Object Access Matrix
- Privilege Lattice and Hierarchical Privileges
- Transmission of rights
- Sharing the data which can check who can access it.
- Role-based Access Control
- Modified communication or VM

Franco and the Agents’ Hell

- A scenario of the worst practices in agents-based software engineering.
- A humoristic story that may be funny for everyone but agent developers.
- A waitress wins with E-shops.
- A musician becomes a beggar because of mistake of an agent.

Agent decisions are tempting.

- Sensors and mechanisms of analyzing data may improve safety.
- Automatization of houses – auto heating, air conditioning, shopping.
- Automatic message and phone answering.
- Agent-based marketplaces with price negotiations suitable for a dynamic economy.

Danger of delegated decision

- Some general security reasoning mechanisms may be tricked.
- Problems with deactivation / update.
- Beer ordering example…
- Agent-based marketplaces with dynamic pricing may dramatically further increase this dynamics.
- Agents systems may work as function of dynamics, not the state.
- Do we trust agents that much (today)?

Conclusions

- Agents creators should work on agents' discipline.
- Study social and political implications of having billions of agents in our physical environment.
- Study and model relationship of the agent system with environment.
- Define modeling tools and methodologies, remember about testing.

Final thoughts…